The Hong Kong University of Science and Technology (Guangzhou)

UG Course Syllabus

[Course Title] Cloud Computing and Big Data Systems

[Course Code] DSAA 4040

[No. of Credits] 3 credits

[Any pre-/co-requisites] UFUG 1601, DSAA 2031

Name: Guoming Tang

Email: guomingtang@hkust-gz.edu.cn

Office Hours: 2:30pm-4:30pm Friday, W3-306

Course Description

Big data systems, including cloud computing platforms and parallel data processing frameworks, have emerged as enabling technologies for managing and analyzing massive volumes of data across hundreds or even thousands of commodity servers in modern datacenters. With the rise of cloud-native infrastructures, such as virtualization, containers, and orchestration platforms, the design and operation of large-scale systems have undergone significant transformation. This course exposes students to both the foundational concepts and hands-on practice of cloud computing and big data systems. The course covers core topics including cloud computing fundamentals and production cloud services, container-based system deployment and orchestration, distributed storage and computing, and modern data processing frameworks. Selected topics on system monitoring, security, and emerging cloud trends are also introduced. By walking through the hands-on lab tasks and exercises, students are expected to gain first-hand experience programming, deploying, and operating systems on real world clusters and cloud environments in production datacenters.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

- Describe the motivation, objectives, and architecture of cloud computing and big data systems.
- Understand the use of production cloud computing platform.
- Understand the general architecture and the use of Hadoop Distributed File System (HDFS).
- Understand the general programming model of MapReduce and the use of Hadoop.
- Understand Resilient Distributed Dataset (RDD) and the use of Spark programming model based on RDD.
- Describe the major architecture difference between Hadoop and Spark.
- Write a MapReduce/Spark program with tens to hundreds lines of code to solve common data analytics problems.
- Use software tools to develop and debug a program written in Hadoop and Spark.
- Understand the concept of serverless computing.
- Understand the infrastructure and software stack of AI cloud.

Weekly schedule & Weekly ILOs

Week	Topics	Weekly ILOs
1	Cloud Computing Concepts and Architecture	ILO 1
2	Virtualization Fundamentals	ILO 1, ILO 2
3	Containerization with Docker	ILO 1, ILO 2
4	Kubernetes Essentials	ILO 1, ILO 2
5	Kubernetes Scaling and Configuration	ILO 2, ILO 3
6	Distributed System Principles	ILO 3, ILO 4
7	Distributed File Systems	ILO 3, ILO 4
8	Apache Spark Core	ILO 5, ILO 6
9	Spark SQL and Structured Streaming	ILO 5, ILO 6, ILO 7
10	Modern Data Architectures	ILO 6, ILO 7
11	System Performance and Observability	ILO 7, ILO 8
12	Cloud Security and Multi-Tenancy	ILO 9, ILO 10
13	Serverless, Edge Computing, and Sustainability	ILO 9, ILO 10

Assessment and Grading

Detailed assessment schemes and grading rubrics are provided below.

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
In-class Quizzes	10%	Regular lecture class
Hands-on Homework	20%	The next lab class
Course Project	20%	Final class week
Mid-term Exam	20%	Week 7 (TBD)
Final Exam	30%	Final exam week (TBD)

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
In-class Quizzes & Homework	ILOs 1-9	The in-class quizzes assess students' understanding of fundamental concepts and system architectures of cloud computing and big data systems (ILO 1), as well as their knowledge of production cloud platforms and distributed storage and computing frameworks such as HDFS, MapReduce, and Spark (ILOs 2-5). The hands-on homework and lab exercises assess students' ability to apply distributed storage and data processing frameworks (ILOs 3-6), develop and execute Spark-based data analytics programs (ILO 7), and use software tools to deploy, monitor, and debug cloud-based applications (ILOs 8-9).

Course Project	ILO 1-10	The course project assesses students' ability to design, deploy, and demonstrate a small cloud or big data system in a realistic environment. It evaluates students' understanding of production cloud platforms and modern cloud-native infrastructures (ILOs 1-2), their ability to apply Spark-based data processing models (ILOs 3-5), analyze architectural trade-offs between different systems (ILOs 6-7), use tools for deployment and debugging (ILO 8), and understand the role of cloud infrastructure in supporting data- and AI-driven applications (ILOs 9-10).
Mid-term Exam	ILOs 1-6	The mid-term exam assesses students' conceptual understanding of cloud computing fundamentals, virtualization, containerization, distributed storage, and parallel data processing models (ILOs 1-5), as well as their ability to compare and explain architectural differences between Hadoop-based and Spark-based systems (ILO 6).
Final Exam	ILOs 1-10	The final exam provides a comprehensive assessment of all intended learning outcomes. It evaluates students' integrated understanding of cloud computing architectures, distributed storage and data processing systems, cloud-native operations, serverless computing, and modern cloud infrastructure, as well as their ability to reason about system design choices and trade-offs (ILOs 1-10).

Grading Rubrics

1. In-class quizzes (10%):

- In-class quizzes are designed to evaluate students' comprehension of the essential subjects. Absences from quizzes will only be excused with a valid medical certificate due to illness.

2. Hands-on Homework (20%):

- Hands-on homework (lab task) must be submitted at the beginning of the next lab class. These hands-on exercises are vital for complementing the concepts discussed in lectures.

3. Group Project (20%):

- Team Formation: It is recommended to form teams of 2~3 members each by the end of Week 2.
- Project Focus: The project may relate to the design, deploy, and demonstrate a small cloud or big data system in a realistic environment and reflect the architectural trade-offs between different systems.
- Project Topic Selection: Each team may choose a topic from the list provided by the course instructor or propose their own topic, subject to the instructor's approval.
- Presentation: Each team is required to deliver a presentation (including Q&A). This presentation is scheduled in the last teaching week.

4. Mid-term and Final Exams (50% in total):

- All exams in this course are closed book and will include a combination of multiple-choice, short-answer, and long-answer questions. These questions are crafted to assess the comprehension of the course material.

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving, and significant creativity in thinking. Exhibits a high capacity for scholarship and collaboration, going beyond core requirements to achieve learning goals.
B	Good Performance	Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn and the ability to work effectively with others.
C	Satisfactory Performance	Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity for analysis and critical thinking. Shows persistence and effort to achieve broadly defined learning goals.
D	Marginal Pass	Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.
F	Fail	Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not meet the threshold requirements for professional practice or development in the discipline.

Course AI Policy

Allowed and encouraged with proper citations and prompt list.

Communication and Feedback

Students can provide informal feedback to the instructor and/or teaching assistant(s) through various channels, including face-to-face conversations, emails, group discussions, or Canvas.

Resubmission Policy

Resubmissions are not permitted after the deadline.

Required Texts and Materials

Not Required. Recommended reading materials will be introduced during classes.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST(GZ)'s Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Regulations for Academic Integrity and Student Conduct for the University's definition of plagiarism and ways to avoid cheating and plagiarism.